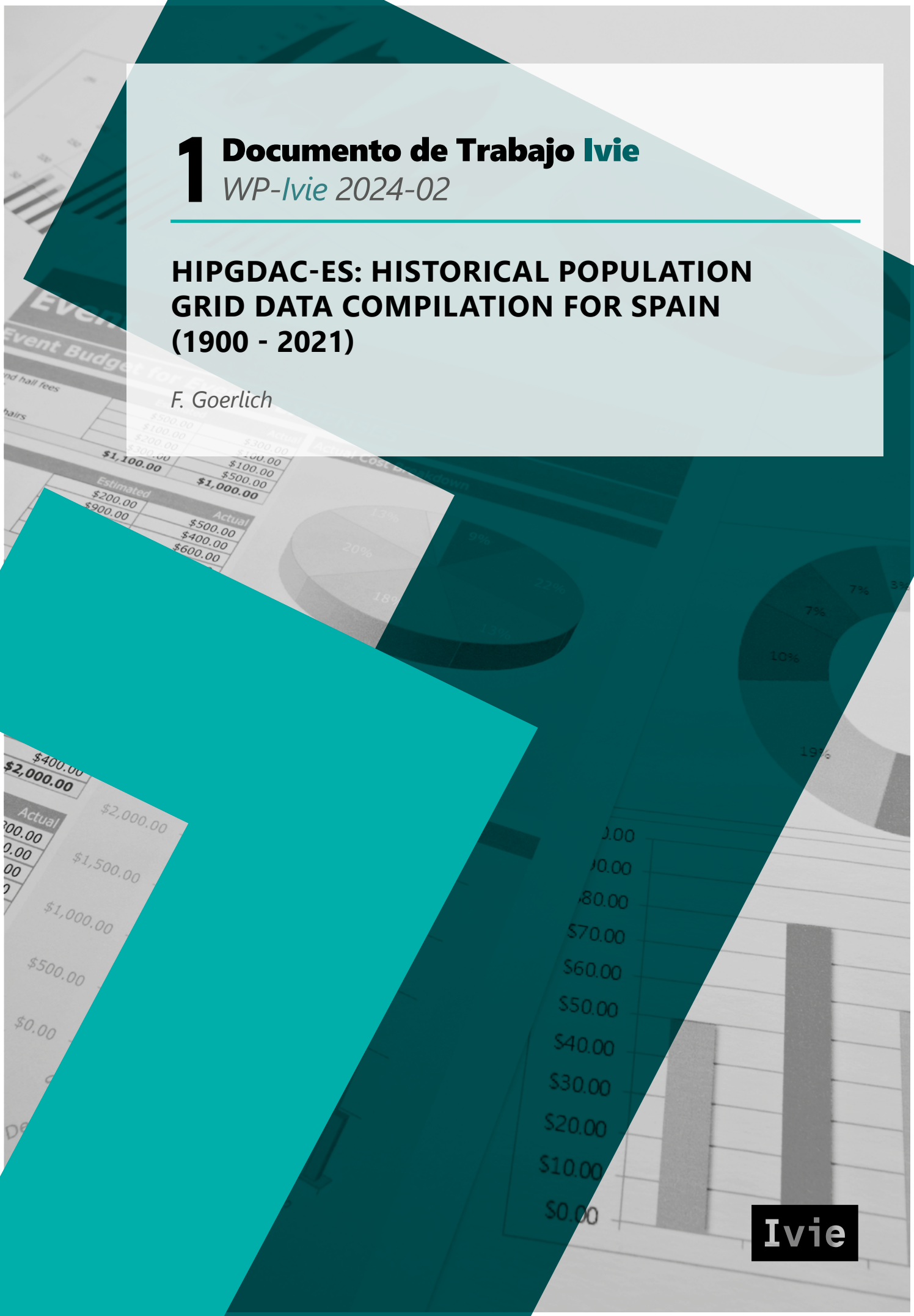


1 **Documento de Trabajo Ivie**

WP-Ivie 2024-02

HIPGDAC-ES: HISTORICAL POPULATION GRID DATA COMPILATION FOR SPAIN (1900 - 2021)

F. Goerlich

Ivie

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso o análisis específicos sobre debates de actualidad, con objeto de divulgar el conocimiento generado por diferentes investigadores.

Ivie working papers offer a preview of the results of economic research under way, as well as an analysis on current debate topics, with the aim of disseminating the knowledge generated by different researchers.

La edición y difusión de los documentos de trabajo del Ivie es una actividad subvencionada por la Generalitat Valenciana, Conselleria de Hacienda y Modelo Económico, en el marco del convenio de colaboración para la promoción y consolidación de las actividades de investigación económica básica y aplicada del Ivie.

The editing and dissemination process of Ivie working papers is funded by the Valencian Regional Government's Ministry for Finance and the Economic Model, through the cooperation agreement signed between both institutions to promote and consolidate the Ivie's basic and applied economic research activities.

Cómo citar/How to cite:
Goerlich, F.J. (2024). « HIPGDAC-ES: Historical Population Grid Data Compilation for Spain (1900-2021) –Version 0 (beta)–». Working Papers Ivie n.º 2024-2. València: Ivie.
http://doi.org/10.12842/WPIVIE_0224

Versión: marzo 2024 / Version: March 2024

Edita / Published by:

Instituto Valenciano de Investigaciones Económicas, S.A.

C/ Guardia Civil, 22 esc. 2 1º - 46020 València (Spain)

DOI: http://doi.org/10.12842/WPIVIE_0224

WP-Ivie 2024-2

HIPGDAC-ES: Historical Population Grid Data Compilation for Spain (1900-2021)

–Version 0 (beta)–

Francisco J. Goerlich[1]

Abstract

Historical population grids are scarce or rather nonexistent. This work represents a first effort in this direction. Using historical cadastral data and homogeneous population data at municipal level, constructed in different projects, we generate, for the whole of Spain, population grids with 100m x 100m and 1km x 1km resolutions and all census years from 1900 to 2021.

These grids are top-down. The methods used are similar to those used in the generation of population grids for the entire globe in the last decades by combining information from satellite images with demographic information from censuses.

Given the richness of cadastral information, and the possibility of going much further back in time than satellite information, we can generate much older population grids. Although far from perfect, these grids provide a better approximation to the distribution of the population in those years than the simple consideration of the municipal population or the count in the settlements derived from the gazetteers associated with the censuses.

The possibility of taking our estimates up to 2021, where we have a bottom-up population grid from the Spanish National Statistical Institute (INE) derived from the 2021 census allows us to validate our methods, albeit only for the most recent dates.

This database should be considered work in progress, subject to validation through experimentation, and may undergo changes in the future.

Keywords: Population; Census; Population Grids; Demography.

JEL classification: J11

Resumen

Las *grids* de población históricas son escasas o más bien inexistentes. Este trabajo representa un primer esfuerzo en esta dirección. Utilizando datos catastrales históricos y una base de datos homogénea de población a nivel municipal, construidos en diferentes proyectos, generamos, para toda España, *grids* de población con resoluciones de 100m x 100m y 1km x 1km y todos los años censales desde 1900 hasta 2021.

Estas *grids* son *top-down*. Los métodos utilizados son similares a los utilizados en la generación de *grids* de población para todo el globo terrestre en las últimas décadas combinando información de imágenes de satélite con información demográfica de censos.

Dada la riqueza de la información catastral, y la posibilidad de retroceder mucho más en el tiempo que la información satelital, podemos generar *grids* de población mucho más antiguas. Aún lejos de ser perfectas, estas *grids* proporcionan una mejor aproximación a la distribución de la población en aquellos años que la simple consideración de la población municipal o el recuento en los asentamientos derivado de los nomenclátores asociados a los censos.

La posibilidad de llevar nuestras estimaciones hasta 2021, donde disponemos de una *grid* de población *bottom-up* procedente del Instituto Nacional de Estadística (INE) derivada del censo de 2021, nos permite validar nuestros métodos, aunque sólo para las fechas más recientes.

Esta base de datos debe considerarse un trabajo en curso, sujeto a validación mediante experimentación, y puede sufrir cambios en el futuro.

Palabras clave: Palabras clave: Población; Censos; *Grids* de población; Demografía.

Clasificación JEL: J11

[1] Universitat de València and Ivie. Contact information: francisco.goerlich@ivie.es.

1.

Introduction

Historical population grids are scarce or rather nonexistent. This work represents a first –but maybe not the last– effort in this direction. Using historical cadastral data and homogeneous population data at municipal level, constructed in different projects, we generate, for the whole of Spain, population grids with 100m x 100m and 1km x 1km resolutions and all census years from 1900 to 2021.

These grids are top-down **(Goerlich and Cantarino 2014).** The methods used are similar to those used in the generation of population grids for the entire globe in the last decades by combining information from satellite images with demographic information from censuses **(Schiavina *et al.* 2023).**

Given the richness of cadastral information, and the possibility of going much further back in time than satellite information, we can generate much older population grids. Although far from perfect, these grids provide a better approximation to the distribution of the population in those years than the simple consideration of the municipal population or the count in the settlements derived from the gazetteers associated with the censuses.

The possibility of taking our estimates up to 2021, where we have a bottom-up population grid from the Spanish **National Statistical Institute (INE)** derived from the 2021 census allows us to validate our methods, albeit for the most recent dates.

The structure of the work is as follows. The data sets used, as well as their advantages and disadvantages, are described in detail below. Section 3 details the methods, and indicates the various products generated. The following section briefly reviews the results obtained for the 1km x 1km grids and the entire period considered. Section 5 performs a tentative validation exercise with the limited information available. Finally, the paper concludes with a brief reflection.

2.

Original information

Although the main sources of information for the generation of the top-down population grids are a historical compilation of cadastre data (Uhl *et al.* 2023) and the homogenised census population data at municipal level –Local Administrative Units (LAU)– (Goerlich *et al.* 2015), the work uses other complementary sources for quality improvement. This section describes in detail the different databases used.

2.1. Historical settlement data compilation for Spain (1900-2020): HISDAC-ES

Uhl *et al.* (2023) download and process the cadastral information of more than 12 million buildings and their features. After a certain harmonisation of the different existing cadastres in Spain –five in total–, they transform the vector information of the building contours into point vector information, and the different features are transformed into raster layers with 100m x 100m resolution. The database is labelled as HISDAC-ES[2]. Validation exercises carried out by Uhl *et al.* (2023) indicate that the database is of acceptable quality, especially in recent times, probably higher than remote sensing products, although at the beginning of the century the quality of the information is considerably reduced due to the problem of dating the buildings in the cadastre, the so-called **survivorship bias**.

The volume of information generated is enormous –743 raster layers– and many of the variables provided have been classified by five-year intervals since 1900. The raster information is provided in different coordinate reference systems (*CRS*), including ETRS89-LAEA[3] –EPSG:3035–, which is the reference system for European grids (INSPIRE 2014), for all Spanish territory.

This information constitutes the main support on which the population will be allocated. Of all the variables offered, the most appropriate for redistributing the resident population by administrative unit over the territory –the grid cells with a resolution of 100m x 100m– is total building indoor area (*RES_BIA*), which is available for each decade from 1900 to 2020 –approximate in line with the decennial census–. This variable includes the two most relevant factors in the processes of spatial disaggregation of the population (Goerlich 2016): the residential

[2] The version used in this work is version 2, which corresponds to the final published work **(Uhl *et al.* 2023)**, and is substantially improved from the initial version. The database is available in ***figshare.***

[3] European Terrestial Reference System 1989 (ETRS89) - Lambert Azimutal Equal Area (LAEA).

function and an indicator of the height of the buildings.

However, cadastral information is not uniform in space and time, and this affects the quality of the population redistribution.

From the spatial point of view, the existence of 5 cadastres for the whole territory, one for the majority of Spain, and 4 for each of the provinces under a different tax system from the rest of the Spanish territory –the 3 provinces of the Basque Country and Navarre– affects the uniformity of the available information.

For example, *RES_BIA* is not available in any year for the 3 Basque provinces –Álava, Guipúzcoa and Vizcaya–. Simply this variable is not in the data model of the cadastres of the Basque Country. In 2020, for Álava we only have information on residential building footprint area (*RES_BUFA*), so that we lose information about the height of the buildings. And for Guipúzcoa and Vizcaya we do not even have information on the use of the building, so that only the building footprint area (*BUFA*) is available. As a result, the disaggregation process should incorporate information from these two variables, *RES_BUFA*, and *BUFA*, in some cases.

As mentioned at the beginning of the section, if we go back in time, things get worse. As recognized by **Uhl *et al.* (2023)**, the HISDAC-ES database has some deficiencies in its early years. This situation is not uniform in space and affects some places more than others. The main problem is the so-called survivorship bias. The HISDAC-ES database infers the age of the building from the date given in the cadastral information, but this date may not correspond to the initial date of construction of the building. The reason is that cadastre changes the date of construction of a building if there has been a restoration or modification of the building of a certain entity, and does not keep record of these changes –at least in the available public information–. As a result, it remains unknown whether a construction date represents the truly first establishment of a building at a given location or whether there was a similar built-up structure existing prior to that. The problem is noted by **Uhl *et al.* (2023)**, section 4.6, which indicate how it is especially evident for some municipalities at the beginning of the 20th century, and urge users to take precautions when analyzing the data in the long term.

There are also minor problems that have to be taken into account. For example, we have the opposite situation to the survivorship bias, buildings that existed in the past, but have been demolished are no longer in the cadastre, hence they are not contained in HISDAC-ES. Also, the function of a cadastre building is the current function, and perhaps in the past its function was different. In addition, it is not certain whether all the characteristics of a building refer to the same date, or if they were measured on the date that is listed as the construction of the building. As already mentioned, HISDAC-ES is the result of combining 5 different cadastres with different data models, and there may be inconsistencies between them. In fact, we have already indicated that some important variables are not available in all cases. Finally, treating buildings as point elements, rather than polygonal, and managing them as raster layers, rather than with the original vector information, introduces certain distortions which, given the magnitude of the database, are necessary in order to reduce the computational burden.

In any case, everything indicates that these are minor problems compared to the survivorship bias, but it is necessary to take them into account when evaluating the results and to consider that these should improve over time, as they have a smaller impact on the currently available information than they did at the beginning of the 20th century.

To get an idea of the importance of the survivorship bias in extreme cases we can examine which municipalities lack cadastral information in the different census years of the twentieth century. If we intersect the municipal boundaries with the layers of *RES_BIA*, *RES_BUFA* and *BUFA* of HISDAC-ES, we will observe that none of the municipalities have zero intersection only from 1970. Even in 1960, there is a municipality that does not have information about buildings in HISDAC-ES. It is a small municipality, Beizama (20020) in the province of Guipúzcoa, with only 460 inhabitants. Of course, this problem gets worse as we go back in time. In 1950, this lack of information affected 4 municipalities, 15 in 1940, 29 in 1930, 57 in 1920, 111 in 1910 and 171 in 1900. In all cases, except for a minor exceptions in the first decades of the 20th century, these are very small municipalities.

It should be noted that this is a general problem that affects all of the municipalities, although it manifests itself in an extreme way in some more than others, and is also present in large cities. Natural aggregation to cells, especially 1km x 1km, dilutes this problem and smooths the survivorship bias, although quantification of the problem is extremely difficult.

A problem related to the survivorship bias, not mentioned in **Uhl *et al.* 2023**, has to do with the dating of many rural isolated farmhouses whose cadastre date does not really correspond to the date of construction. Many of these buildings were inhabited at the beginning of the 20th century and gradually ceased to be so throughout the second half of that century. The indications of this incorrect dating are evident from the fieldwork and direct consultation to the cadastre data of certain buildings. The problem in some places is so obvious that a direct and personal consultation on this matter was made to the rural section of the cadastre. The answer is particularly revealing:

"…it must be taken into account that the collection of information on the dates of construction of buildings on rural land, was carried out for most of the national territory in a massive way in the 1990s.

This work was contracted to external companies which made a construction sheet that reflected among other features, its use and an estimated date of seniority.

The nature of the fieldwork carried out, as well as the enormous task of covering the entire national territory, sometimes prevented the exact date of construction from being known, in such a way that it is not possible to guarantee the usefulness of this data to be used in publications of a scientific nature that may require more rigour." (e-mail dated 28/04/2023)[4]

There is no apparent solution to this type of problem, but it is necessary to be aware of it

[4] The consultation was carried out through Luis Julián Santos Pérez, Chief of Service of the Directorate General of Cadastre, and specialist in the historical cartography of Cadastre **(Santos 2019).**

in order to be able to evaluate the results obtained.

What is obvious is that the information of HISDAC-ES must be complemented with another, especially in the first decades of the 20th century, if we start from municipal information to generate historical grids of population.

2.2. Historical municipal population data: 1900-2021

The main source of population information is the homogeneous database of **Goerlich, Ruiz, Chorén and Albert (2015).** These authors generate estimates of municipal population with the structure of municipalities from the 2011 census, so that they take into account the municipal alterations –mergers and segregations, total and partial– occurred between 1900 and 2011. We therefore have homogeneous municipal populations for the 8,116 municipalities that appear in the 2011 census. In addition, to reach the most recent date, the database is expanded with the populations of the 8,131 municipalities of the 2021 census.

The contours of the municipalities come from the **National Geographical Institute (IGN)** with reference dates 01/01/2012, for the period 1900-2011 –8,116 municipalities–, and 01/01/2019, for 2021 –8,131 municipalities–. Since the **IGN** municipal enclosures and boundary limit lines database is continuously updated and the **IGN** does not maintain historical data, these are collected and organized by **Goerlich and Perez (2021).**

2.3. Settlement coordinates from 1887 census: *ESPAREL*

As has become clear in the description of HISDAC-ES at the beginning of this section, we must incorporate external information that mitigates, as far as possible, the survivorship bias.

An independent geocoding project of the historical population entities contained in the gazetteers of the late eighteenth and nineteenth centuries is **ESPAREL**[5]. The main objective of **ESPAREL** has been to generate a spatial data infrastructure (SDI) that allows linking the spatial planning of the Old Regime with that of the liberal State at the end of the 19th century, and at the same time with the current one. For this purpose, the existing population entities in the 1787 census –census of Floridablanca–, and those of the 1887 Spanish Gazetteer –corresponding to the census of the same year– have been linked with the *Basic General Gazetteer of Spain* (NGBE), which is geocoded, as well as other external geocoded information currently available from regional cartographic institutes or map viewers accessible through external queries from APIs.

From our point of view, what matters is that the gazetteer of 1887 is relatively close to 1900, which is our first year for the generation of historical population grids from HISDAC-ES, and for which we have the point coordinates of 58,837 population entities, covering almost all of the current municipalities. We incorporate this information for those municipalities in which HISDAC-ES lacks it. This solves all the cases mentioned above

[5] The **ESPAREL** project has been financed by the **BBVA Foundation** within the Program of '*Aid to Scientific Research Teams in Digital Humanities*', call 2019.

except those of 4 newly created municipalities, for which the homogeneous populations of **Goerlich, Ruiz, Chorén and Albert (2015)** indicate the existence of population, but not cadastre –HISDAC-ES–, neither **ESPAREL** gives geocoded information that allows to bound in the space the population of these municipalities[6].

It should be remembered that the geocoded information of **ESPAREL** does not refer to buildings, but to the population entity itself. However, the number of buildings and their plants are available in the database and, by construction, they are residential buildings. From this information a variable was constructed which adds the buildings once multiplied by their plants, which in a certain way represents the closest to the variable *RES_BIA* of HISDAC-ES, except that the coordinates do not correspond to individual buildings, but to the population entity. This vector point information was transformed into a raster layer with the same structure as the HISDAC-ES layers. Note that each entity will belong to a single cell of 100m x 100m, regardless of its size, but since these are entities of low demographic weight this is not expected to generate many distortions, especially on cells added to a resolution of 1km x 1km.

As a last resort, for the municipalities for which we do not find information either in **ESPAREL**, we use the coordinate of the capital of the municipality, which is available in the *Gazetteer of Municipalities and Population Entities* (NGMEP) of the **IGN**. Like the coordinates of **ESPAREL** this vector point information was transformed into raster layers with the same structure as those of HISDAC-ES, and the population of the municipality assigned to the corresponding cell. The small size of the population in this case makes distortions minimal.

2.4. Other information used

In addition to the above information, used in the process of downscaling the municipal population to the grid, the **Global Human Settlement Layer (GHSL) grid population data (Schiavina, Freire, Carioli and MacManus 2023)**, version R2023A **(Schiavina *et al.* 2023)**, with resolution of 1km x 1km, were used for the years 1980, 1990, 2000, 2010 and 2020 to compare their results with those obtained by our methods. The **GHSL** is a worldwide product that had to be masked by Spain's administrative boundaries for comparison with our results.

[6] These municipalities are Badia del Vallès (08904), Vegaviana (10902), Alagón del Río (10903) and Vencillón (22909), and its population only exceeds 1000 inhabitants for Badía del Vallès (08904) in 1950. In the other years the municipalities for which we lack support to allocate the population are extremely small. For example, in 1900 these populations ranged between 276 inhabitants for Badía del Vallès (08904) and 5 inhabitants for Alagón del Río (10903).

3.

Downscaling historical municipal population data: Methods

The methods are relatively simple, and are not very different from those used in the production of the GHSL population grids **(Freire, MacManus, Pesaresi, Doxsey-Whitfield and Mills 2016).**

The approach is based on the well-known raster-based dasymetric mapping (Wright 1936, **Mennis 2003)** relying mainly on the variable *RES_BIA* from HISDAC-ES to restrict and refine the distribution of people. Even in 1900, this covers 93% of municipalities, that increases to 97% in 2021. When, for a given municipality, this information is insufficient or non-existent other less appropriate variables are used following a clear predefined order.

For each period, population grids were produced following a clear and concise volume-preserving dasymetric mapping approach at LAU level. So, for a given period we iterate over municipalities. Given a vector census layer of municipal boundaries and a HISDAC-ES raster layer, for a populated municipal polygon (source zone) we proceed as follows:

1. If a particular municipality has information on the **total building indoor area** (*RES_BIA*), then we downscale population proportionally to *RES_BIA*.

2. If there is no information on *RES_BIA*, but the municipality has **residential building footprint area** (*RES_BUFA*), then we downscale population proportionally to *RES_BUFA*.

3. If there is no information on the previous variables, but the municipality has **building footprint area** (*BUFA*), then we downscale population proportionally to *BUFA*.

4. If there is no information on HISDAC-ES for a given municipality, then we use the **ESPAREL** coordinates of population entities to redistribute the municipal population.

5. As a last resort, if no other information is found in neither HISDAC-ES or **ESPAREL** we assigned municipal population to the coordinates of the municipal capital.

It should be obvious that the process is carried out for each year and municipality sequentially.

The raster dasymetric mapping approach is implemented on the original resolution of the HISDAC-ES raster layers, 100m x 100m, and the population per cell rounded to integer values at this resolution using LAU population as totals –with the exception of the 2011 census, ¡whose population values published by **INE** are real, 👎!–. Hence, population grids are integer valued at this resolution, and aggregated afterwards to the standard resolution of 1km x 1km, in raster and vector formats.

Table 1, below, shows the population of each census and the number of inhabited cells estimated with the above algorithm. According to our estimates, while the population has multiplied by 2.5 in the period 1900-2021, the inhabited cells have only doubled.

Table 1. Census population and estimated inhabited grid cells

Year	Population	Cells
1900	18,830,649	71,767
1910	20,360,306	74,151
1920	22,012,663	80,511
1930	24,026,571	86,32
1940	26,386,854	93,146
1950	28,172,268	101,09
1960	30,776,935	108,576
1970	34,041,482	116,357
1981	37,682,355	127,34
1991	38,872,268	133,553
2001	40,847,371	138,636
2011	46,815,916	148,719
2021	47,400,798	143,292

Source: HISDAC-ES and Census 1900-2021.

A close inspection of table 1 reveals some striking figures that deserve a few comments. The number of inhabited cells continuously increases up until the last two years, 2011 and 2021. In this case the last census, 2021, shows a significant drop in the number of inhabited cells compared to the previous census, 2011. One might think that this trend is due to the huge population growth between 2001 and 2011, as opposed to the low population growth between 2011 and 2021. However, this is not the case. It is easy to show that the result is, in a way, a statistical artifact derived from the fact that the **INE** published population data in real figures for the 2011 census, which as **INE (2022)** itself acknowledges was not a good idea, and introduced additional complications. Our estimates maintain the INE population structure, and therefore the layers generated provide population figures in integers at the 100m x 100m cell level for all years except 2011, where the results of the disaggregation process are not rounded. The result is that the population in real figures overestimates the number of inhabited cells. If for 2011 we round off the population figures to integers at the 1km x 1km cell level[7], then the population obtained is 46,815,731 persons, slightly different from the total census population –table 1–, and the number of inhabited cells would be 143,808, much more in line with the trend observed in table 1. Even in this case the number of inhabited cells would be slightly lower in 2021 than in 2011, but given the magnitude of the differences we cannot rule out that this would not be the case if the rounding were done at the level of 100m x 100m cells.

Another curious fact, which is not possible to observe only with the information in table 1, is that the trend observed for the years 2011 and 2021 in our estimates is radically different from that observed when we compare the only two population grids published by **INE**, derived by bottom-up procedures, and which correspond to the 2011 and 2021 censuses, where the inhabited cells in 2021 practically double those that existed in 2011 **(Goerlich 2024).** As indicated in **Goerlich**

[7] These results are not invariant to the geographical scale at which we round to integers, so the numbers would be different if we rounded at the 100m x 100m cell level.

(2024), section 5, this is due to a design effect in the generation of the 2011 grid, which makes it unreliable for determining the actual number of inhabited cells in Spain in that year.

As a result of this process, we have 13 raster population layers with resolution of 100m x 100m, 13 raster population layers with resolution of 1km x 1km, and a vector file with all populated cells in at least one census year –geopackage format (*gpkg*)–, as well as the cell indicator according to the **INSPIRE (2014)** directive[8]. This information can be accessed from the following .

The whole process was performed using free software based on the statistical computing system *R* **(R Core Team 2023),** using ***tidyverse* libraries** (Wickham et al. 2019) for data wrangling, ***sf*** library (Pebesma 2018) for handling vector information and ***terra*** (Hijmans 2023) and ***stars*** (Pebesma and Bivand 2023) libraries for handling raster information.

[8] This information can be accessed from the following **site:** https://zenodo.org/records/10818016.

4.

Historical population grid data at 1km x 1km resolution: 1900-2021

This section illustrates some of the results obtained from the 1km x 1km resolution population grids. In a way, it has a dual purpose. On the one hand, it is a first descriptive analysis of the historical evolution of the population distribution in this format. On the other hand, it serves as a first tentative validation exercise, in the sense of examining whether the results obtained are in agreement with the main known historical trends in the evolution of the distribution of the Spanish population throughout the 20th century and the beginning of the 21st century. These trends are well known **(Goerlich and Molla 2021).**

First, table 2 shows the distribution of cell sizes in relative terms over the period 1900-2021[9]. Large trends are observed at the extremes of the distribution. Thus, cells with a small population, up to 10 inhabitants, go from representing 12% in 1900 to more than tripling in 2021, representing 42%. At the other extreme, cells with more than 1,000 inhabitants also increase their relative weight, although only those with more than 10,000 inhabitants show a marked trend. In this case, their relative importance is low, not even reaching 1% in 2021, but their weight has increased fivefold in these 120 years. The remaining cell sizes, above 25 and up to 1,000 inhabitants, show a decreasing relative importance, especially in the interval between 100 and 200 inhabitants.

Second, table 3 shows the percentage of inhabited cells out of the total at the Autonomous Community level, i.e. the proportion of inhabited area with a resolution of 1km x 1km. Naturally, given the increase in population, all regions show increasing percentages of human occupation, but in some regions the trend is particularly striking. Thus, the Community of Madrid multiplies its occupied area by 6.4, Andalusia by 4.2, Extremadura by 4.8 and the Region of Murcia by 2.5. The results also show, clearly, the regions with the most dispersed population –in the sense of showing greater human occupation of their territory–, Galicia, Asturias and the Basque Country. In 2021, with the exception of Ceuta and Melilla, only Galicia exceeds 60% of populated territory, although the Balearic Islands, the Basque Country and the Principality of Asturias exceed 50%, whereas the Region of Murcia is right in this percentage.

Two well-known trends in the evolution of population distribution throughout the 20th century is its tendency to move towards the valley and towards the coast **(Goerlich and Molla 2021)**[10]. Using data from **Goerlich (2023)** it is possible to quantify these trends from the generated population grids.

[9] Given the population increase over the period 1900-2021 all results are shown in relative terms.

[10] The rest of the trends, the displacement towards urban municipalities and provincial capitals **(Goerlich and Molla 2021)**, cannot be examined on the basis of the generated information, since the origin of it is municipal data.

Table 2. Distribution of cells (%) by population size

Interval	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011	2021
Up to 10	12.05	11.97	13.06	14.52	16.58	19.58	23.24	29.17	34.83	37.81	40.05	42.64	42.25
(10, 25]	14.96	14.79	15.45	15.95	16.38	16.89	17.07	17.00	16.85	16.70	16.30	15.85	15.90
(25, 50]	15.58	15.27	14.92	14.92	14.75	14.51	14.06	13.75	12.96	12.38	12.00	11.15	11.13
(50, 100]	17.16	16.93	16.90	16.40	15.83	15.05	14.02	12.89	11.55	10.77	9.91	8.96	8.91
(100, 200]	16.05	16.10	15.79	15.15	14.35	13.28	12.27	10.53	9.03	8.22	7.54	6.76	6.65
(200, 300]	7.51	7.72	7.43	6.95	6.56	6.06	5.42	4.45	3.82	3.44	3.21	3.05	3.08
(300, 500]	6.69	6.86	6.45	6.29	5.90	5.40	4.95	4.13	3.54	3.26	3.18	3.08	3.14
(500, 1000]	5.45	5.50	5.24	5.01	4.90	4.58	4.22	3.57	3.12	3.06	3.06	3.13	3.25
(1000, 5000]	4.08	4.35	4.23	4.21	4.08	3.95	3.96	3.55	3.25	3.30	3.59	4.05	4.27
(5000, 10000]	0.34	0.35	0.37	0.40	0.43	0.46	0.50	0.51	0.52	0.53	0.60	0.75	0.80
(10000, 20000]	0.08	0.10	0.10	0.13	0.14	0.14	0.18	0.28	0.31	0.32	0.36	0.41	0.44
More than 20000	0.05	0.05	0.06	0.07	0.09	0.09	0.12	0.16	0.22	0.21	0.19	0.18	0.19
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Source: HISDAC-ES and Census 1900-2021.

Table 4 shows the evolution of population distribution by altimetric cuts for the period 1900-2021, where altitude is measured, as is population, at cell level **(Goerlich 2022).** For comparability, the same altimetric cuts are used as in **Goerlich and Molla (2021)**, and broadly very similar trends are observed. If the resident population below the 200 meters range was about a third in 1900, in 2021 it reached 53%. The rest of the intervals decreased their relative importance, with the strip above 1,000 meters of altitude showing the biggest decrease.

Finally, Table 5 shows the evolution in the distribution of the population according to its distance to the coast in different intervals[11]. Distance, like population, is measured at cell level **(Goerlich 2023).**

Clearly, Spain was already a coastal country in 1900 **(Rappaport and Sachs 2003),** since a quarter of its population was located on a 10km strip from the coastline, but in 2021 this percentage reached 40%. The rest of the intervals, between 10 and 200km, have seen their relative population decrease, and the slight rise in population observed at more than 200km off the coastline, 3 percentage points, is undoubtedly due to the increase of the metropolitan area of Madrid **(Goerlich and Molla 2021).**

On the whole, the results seem reasonable and show the major trends already known about the evolution in population distribution throughout the 20th and early 21st century, measured at the municipal level **(Goerlich and Molla 2021),** although we are far from a real validation exercise, but at least this shows that the estimated grids makes sense.

[11] Distances are measured in a straight line from the cell to the coastline of the layer of administrative contours of **IGN (Goerlich and Perez 2021).** Calculations were made on the grid projection, **ETRS89-LAEA.**

Table 3. Share (%) of inhabited cells by Autonomous Community

Code	Region	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011	2021
1	Andalucía	7.2	7.9	9.7	11.8	14.5	17.8	20.5	22.6	25.2	26.7	28.3	30.9	30.0
2	Aragón	5.8	6.0	6.3	6.6	7.1	7.7	8.2	8.8	10.0	10.6	10.8	12.0	11.1
3	Principado de Asturias	47.6	48.2	48.8	49.2	49.6	50.5	51.6	51.9	52.1	52.2	52.2	53.9	51.8
4	Illes Balears	31.6	32.2	36.0	37.5	39.5	41.1	43.1	46.6	50.0	51.8	53.2	54.9	54.6
5	Canarias	21.8	22.6	24.3	25.9	27.7	30.3	33.6	36.4	40.1	42.7	45.8	47.8	47.7
6	Cantabria	25.8	26.4	34.0	38.0	39.0	39.7	40.2	41.0	41.6	42.3	42.5	43.6	42.8
7	Castilla y León	9.0	9.3	10.3	10.8	11.3	11.8	12.3	12.8	13.9	14.6	15.0	16.2	15.3
8	Castilla-La Mancha	4.3	4.5	5.1	5.7	6.4	7.2	8.0	9.0	11.1	12.4	13.7	15.9	14.7
9	Cataluña	30.8	31.6	32.5	33.4	34.4	36.2	37.9	40.5	43.5	45.1	46.1	48.3	47.2
10	Comunidad Valenciana	19.7	20.8	23.0	25.6	27.9	30.4	33.6	38.3	43.4	45.6	46.7	50.1	47.5
11	Extremadura	3.7	3.9	4.7	5.7	7.8	9.6	11.3	12.5	14.3	15.4	16.6	19.5	17.8
12	Galicia	54.6	55.6	57.4	58.7	59.7	60.5	61.2	62.0	63.1	64.1	64.6	65.7	64.8
13	Comunidad de Madrid	6.4	6.8	7.4	8.5	10.2	12.6	17.0	23.6	33.0	36.1	38.2	41.3	41.2
14	Región de Murcia	20.2	21.7	26.1	29.5	32.8	38.0	40.9	43.1	45.6	47.2	48.4	50.3	50.0
15	Comunidad Foral de Navarra	11.0	11.1	11.3	11.5	11.8	12.2	12.8	13.6	15.0	15.9	16.7	17.7	17.8
16	País Vasco	35.4	34.8	36.0	37.3	38.7	40.4	41.5	44.0	47.8	49.5	51.1	52.7	52.4
17	La Rioja	8.7	8.9	9.3	9.6	10.2	11.1	12.0	13.8	15.7	16.7	17.2	19.3	17.5
18	Ceuta	15.4	15.4	23.1	35.9	41.0	51.3	59.0	59.0	64.1	64.1	64.1	64.1	64.1
19	Melilla	11.1	19.4	30.6	36.1	36.1	38.9	41.7	41.7	41.7	41.7	47.2	47.2	47.2
	Total	**13.9**	**14.4**	**15.6**	**16.8**	**18.1**	**19.6**	**21.1**	**22.6**	**24.7**	**25.9**	**26.9**	**28.9**	**27.8**

Source: HISDAC-ES and Census 1900-2021.

Table 4. Population share (%) by altitude (m)

Altitude	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011	2021
Up to 200m	33.5	34.1	35.2	36.3	37.8	39.1	41.7	46.5	49.8	50.9	51.6	52.4	53.0
(200, 600]	30.7	30.5	30.1	29.4	28.6	27.9	26.9	24.7	23.2	22.7	22.5	22.3	22.0
(600, 1000]	30.2	29.9	29.7	29.6	29.4	29.0	27.9	26.4	25.2	24.7	24.4	23.9	23.7
More than 1000m	5.7	5.4	5.1	4.7	4.3	4.0	3.5	2.4	1.8	1.6	1.5	1.4	1.3
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: HISDAC-ES and Census 1900-2021.

Table 5. Population share (%) by altitude (m)

Distance	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011	2021
Up to 10km	24.5	25.1	25.7	26.8	27.9	28.8	30.6	34.7	37.2	37.8	38.2	39.0	39.5
(10, 50]	22.8	22.3	21.7	20.8	20.0	19.4	19.2	18.6	18.4	18.3	18.4	18.5	18.6
(50, 100]	15.8	15.7	15.6	15.2	15.0	14.8	13.9	12.5	11.7	11.6	11.3	10.9	10.7
(100, 200]	19.6	19.4	19.3	19.1	18.8	18.4	17.1	14.5	13.1	12.7	12.1	11.3	10.7
More than 200km	17.4	17.5	17.7	18.1	18.2	18.6	19.2	19.6	19.6	19.7	20.0	20.3	20.5
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: HISDAC-ES and Census 1900-2021.

5.

Comparison with other population grids

5.1. Global Human Settlement Layer Population grids

The **Global Human Settlement Layer (GHSL)** publishes, in its version R2023A, population grids between 1975 and 2030 with a five-year periodicity[12], resolutions of 100m x 100m and 1km x 1km and projection Mollweide[13]. We can therefore compare our results for the years 1981, 1991, 2001, 2011 and 2021 with the corresponding GHSL results for 1980, 1990, 2000, 2010 and 2020. After all, our disaggregation methods are similar to those used by the **GHSL (Freire, MacManus, Pesaresi, Doxsey-Whitfield and Mills 2016),** although our support where to place the population differs substantially. In our case the background information comes mostly from the Cadastre, while in the case of the **GHSL** the information comes mainly from raster layers of built-up area with a functional classification –residential versus non-residential– and an estimation of built-up volume through building height.

The comparison is made only from the 1km x 1km resolution. Once the tessellations corresponding to Spain have been downloaded and extracted, we join them together and cut them by the IGN administrative contours identical to those used for the generation of the population grid, once all the information has been transformed to the same coordinate reference system (*CRS*). The **GHSL** provides population figures per cell in real terms –and we have already observed that this is not a trivial matter when determining the number of inhabited cells–. To be consistent with our estimates we round cell-level population figures to integers, with the exception of 2010, as the 2011 census population figures are real, so we must retain this feature for comparison purposes.

Table 6 compares the population and inhabited cells from our estimates –table 1– with what we get from the **GHSL** from the previous process. The populations are reasonably similar, but the number of inhabited cells is not. The **GHSL** shows a much higher number of inhabited cells than our estimates, in the order of twice as many inhabited cells. All indications are that **GHSL** estimates disperse the population excessively. For 2011 we observe, in the **GHSL** estimates, the same effect as in HIPGDAC-ES related to the actual population. In fact, if we round up the **GHSL** population in 2011 at cell level the number of inhabited cells decreases to 265,081, slightly above 10%.

[12] Of course, grids for the years 2025 and 2030 are projections.

[13] The information is also provided in geographic coordinates, **WGS84** reference system, with resolutions of 3 and 30 arcseconds.

Table 6. Population and estimated inhabited grid cells: HIPGDAC-ES *versus* GHSL

HIPGDAC-ES			GHSL		
Year	Population	Cells	Year	Population	Cells
1981	37,682,355	127,34	1980	37,539,291	261,795
1991	38,872,268	133,553	1990	38,941,450	263,393
2001	40,847,371	138,636	2000	40,796,668	261,898
2011	46,815,916	148,719	2010	46,634,491	298,11
2021	47,400,798	143,292	2020	47,424,225	262,423

Source: HIPGDAC-ES, Census 1900-2021 and GHSL2023A.

Table 7 displays the distribution according to cell sizes in relative terms during the years that are common between HIPGDAC-ES and **GHSL**. The differences here are quite remarkable. **GHSL** estimates do not show a definite trend, but a high stability, unlike the estimates of HIPGDAC-ES. The proportion of cells of very small size, up to 10 inhabitants, is significantly higher in **GHSL** than in HIPGDAC-ES, and exceeds 50% in all years. The opposite is true for larger cells, where **GHSL** always offers less relative importance. This minor relative importance is actually perceptible in all cell sizes above 25 inhabitants.

To compare the similarity between percentage structures we can calculate the following discrepancy statistic:

$$\delta = \frac{1}{2}\sum_j |S_{1j} - S_{2j}| \qquad [1]$$

Where S_1 and S_2 represent the percentage structures under comparison.

This statistic, δ, ranges between 0, if both percentage structures are identical, and 1, in the case of maximum discrepancy, when percentage structures do not overlap.

The value of δ, expressed in percentage terms, $100 \times \delta$, ranges from 17.2 in 1981 to 9.8 in 2021, with a monotonous downward trend. This convergence is clearly due to the tendency observed in the evolution of the percentages of HIPGDAC-ES, showing a clear growth in time of the smaller cells, which generates an approximation of the percentage structure of HIPGDAC-ES to the percentage structure of **GHSL.**

Together, our estimates, HIPGDAC-ES, and those of **GHSL** show notable differences in the few dimensions in which we have compared them. What is not clear is which of the two estimates is more realistic, as everything indicates that **GHSL** has to disperse the population in excess, with a higher relative number of cells than would be reasonable for the smaller cell size.

Table 7. Distribution of cells (%) by population size: HIPGDAC-ES *versus* GHSL

	HIPGDAC-ES					GHSL				
Interval	**1981**	**1991**	**2001**	**2011**	**2021**	**1980**	**1990**	**2000**	**2010**	**2020**
Up to 10	34.83	37.81	40.05	42.64	42.25	51.64	51.04	51.85	54.51	51.48
(10, 25]	16.85	16.70	16.30	15.85	15.90	17.19	17.37	16.82	15.50	16.47
(25, 50]	12.96	12.38	12.00	11.15	11.13	10.76	10.92	10.52	9.64	10.19
(50, 100]	11.55	10.77	9.91	8.96	8.91	8.17	8.20	7.87	7.35	7.73
(100, 200]	9.03	8.22	7.54	6.76	6.65	5.14	5.17	5.06	4.92	5.16
(200, 300]	3.82	3.44	3.21	3.05	3.08	1.88	1.90	1.92	1.94	2.09
(300, 500]	3.54	3.26	3.18	3.08	3.14	1.51	1.56	1.66	1.74	1.93
(500, 1000]	3.12	3.06	3.06	3.13	3.25	1.42	1.45	1.65	1.67	1.87
(1000, 5000]	3.25	3.30	3.59	4.05	4.27	1.72	1.79	2.03	2.11	2.36
(5000, 10000]	0.52	0.53	0.60	0.75	0.80	0.32	0.33	0.36	0.36	0.40
(10000, 20000]	0.31	0.32	0.36	0.41	0.44	0.19	0.19	0.21	0.20	0.23
More than 20000	0.22	0.21	0.19	0.18	0.19	0.06	0.07	0.06	0.06	0.09
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Source: HIPGDAC-ES, Census 1900-2021 and GHSL2023A.

5.2. GEOSTAT2021: Census 2021 population grid

Finally, we make a brief comparison of our estimates for 2021, HIPGDAC-ES, with the grid generated by **INE** from the 2021 census, GEOSTAT2021, which has been generated by bottom-up methods and, in principle, we consider it to be the reference grid[14].

The first difference between our estimates and GEOSTAT2021 refers to the number of inhabited cells. While GEOSTAT2021 indicates that we have 115,410 inhabited cells, our estimate gives a value of 143,292 cells –table 1–, i.e. 24% more. This gives us a first indication that HIPGDAC-ES tends to show a greater dispersion of the population than the real one, understanding by this the one offered by GEOSTAT2021.

Table 8 compares the distribution of cells, in relative terms, by population size for 2021 resulting from our estimates, HIPGDAC-ES, with that observed in GEOSTAT2021. This percentage distribution is relatively similar, and discrepancies are only observed at the extremes of the distribution. GEOSTAT2021 shows a lower percentage of very small cells than that observed in HIPGDAC-ES, up to 10 inhabitants, and a higher percentage of larger cells. The discrepancy index, $100 \times \delta$, between both percentage structures is 2.9, much lower than the one we observed when comparing HIPGDAC-ES with the **GHSL**, which was 9.8 for the same year.

[14] The reasons why we did not make the same comparison with the grid generated by INE from the 2011 census, GEOSTAT2011, is because its production method is not comparable with GEOSTAT2021 and generated some undesirable results **(Goerlich (2024), section 5)**.

Table 9 compares the distribution of inhabited cells by Autonomous Community and, in this case, the differences are notable, as a result of the fact that HIPGDAC-ES offers a greater number of inhabited cells. In fact, in all regions, except in the Balearic Islands, Navarre, Ceuta and Melilla, GEOSTAT2021 offers a lower percentage of inhabited cells than HIPGDAC-ES, the differences being particularly striking in the Region of Murcia and the Community of Valencia. At the national level, while HIPGDAC-ES indicates that inhabited cells account for 27.8% of the total, GEOSTAT2021 reduces this percentage to 22.4%.

Table 10 compares the percentage distribution in terms of the altimetric cuts and shows a high level of agreement. The discrepancy index, in percentage terms, $100 \times \delta$, is only 0.28.

Table 8. Distribution of cells (%) by population size: HIPGDAC-ES *versus* GEOSTAT2021

Interval	HIPGDAC-ES	GEOSTAT2021
Up to 10	42.25	39.35
(10, 25]	15.90	15.89
(25, 50]	11.13	11.42
(50, 100]	8.91	9.39
(100, 200]	6.65	6.88
(200, 300]	3.08	3.31
(300, 500]	3.14	3.38
(500, 1000]	3.25	3.53
(1000, 5000]	4.27	4.93
(5000, 10000]	0.80	1.04
(10000, 20000]	0.44	0.62
More than 20000	0.19	0.25
Total	**100.00**	**100.00**

Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

Table 9. Share (%) of inhabited cells by Autonomous Community: HIPGDAC-ES *versus* GEOSTAT2021

Code	Region	HIPGDAC-ES	GEOSTAT2021
1	Andalucía	30.0	21.6
2	Aragón	11.1	7.9
3	Principado de Asturias	51.8	48.9
4	Illes Balears	54.6	64.9
5	Canarias	47.7	46.5
6	Cantabria	42.8	41.1
7	Castilla y León	15.3	13.0
8	Castilla-La Mancha	14.7	7.8
9	Cataluña	47.2	40.2
10	Comunidad Valenciana	47.5	36.2
11	Extremadura	17.8	9.6
12	Galicia	64.8	62.5
13	Comunidad de Madrid	41.2	35.3
14	Región de Murcia	50.0	35.9
15	Comunidad Foral de Navarra	17.8	18.8
16	País Vasco	52.4	50.1
17	La Rioja	17.5	13.1
18	Ceuta	64.1	69.2
19	Melilla	47.2	50.0
	Total	**27.8**	**22.4**

Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

Table 10. Population share (%) by altitude (m): HIPGDAC-ES *versus* GEOSTAT2021

Altitude	HIPGDAC-ES	GEOSTAT2021
Up to 200m	53.0	53.3
(200, 600]	22.0	21.9
(600, 1000]	23.7	23.6
More than 1000m	1.3	1.2
Total	**100.0**	**100.0**

Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

Finally, table 11 compares the percentage distribution in terms of distance to the coast. In this case the percentages in the various cutoffs are virtually identical. The discrepancy index, in percentage terms, $100 \times \delta$, is only 0.12.

Overall, the results show a high degree of agreement in relative terms, although significant discrepancies remain in absolute terms, i.e. in terms of inhabited cells.

Table 11. Population share (%) by distance to the coast (km): HIPGDAC-ES *versus* GEOSTAT2021

Distance	HIPGDAC-ES	GEOSTAT2021
Up to 10km	39.5	39.4
(10, 50]	18.6	18.7
(50, 100]	10.7	10.7
(100, 200]	10.7	10.8
More than 200km	20.5	20.5
Total	**100.0**	**100.0**

Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

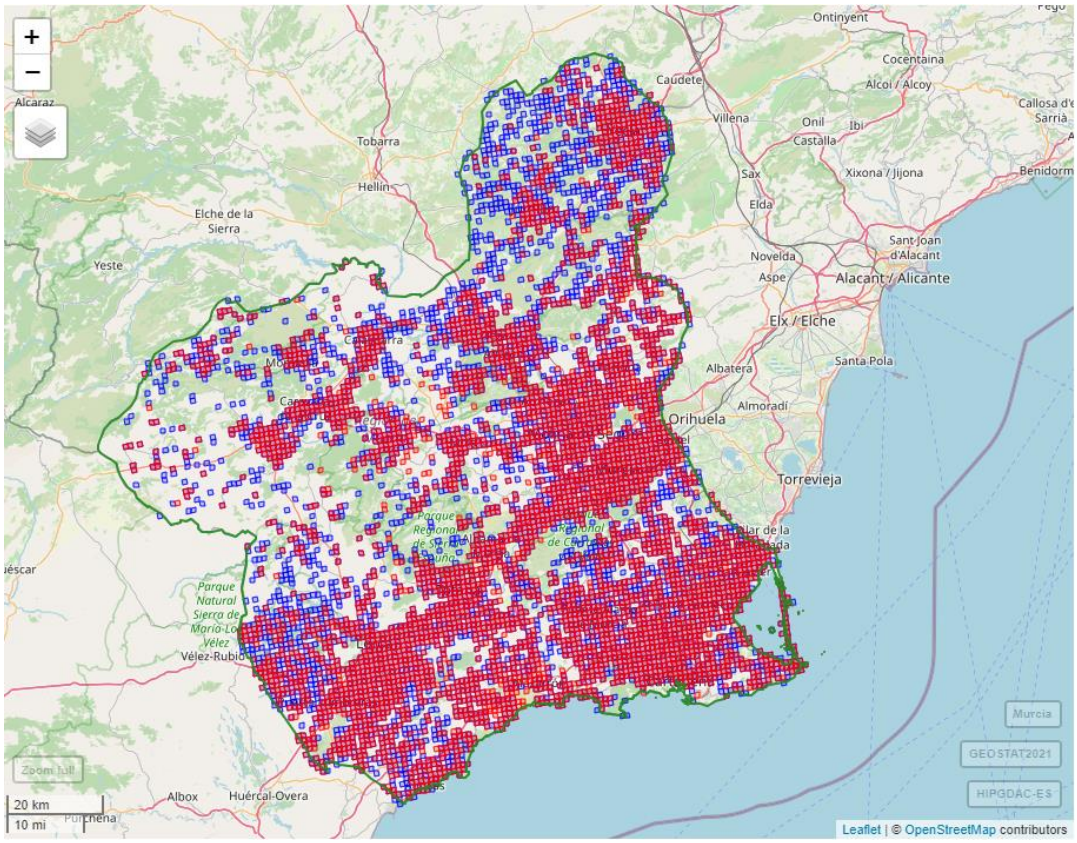
It is illustrative to calculate the confusion matrix –crosstabulation– between the two grids for 2021. This matrix is given in both, absolute and relative terms in table 12. Of the inhabited cells in GEOSTAT2021 only 8.3% –9,598 cells– are not inhabited in HIPGDAC-ES. Thus, almost all the inhabited cells in the INE census grid are also inhabited in our estimates. On the contrary, of the inhabited cells in HIPGDAC-ES, there are 26.2% of cells –37,480 cells– that are not inhabited in GEOSTAT2021.

Table 12. Cross tabulation of cells: HIPGDAC-ES *versus* GEOSTAT2021

	2021					
	Cells			Percentages (%)		
	YES	NO	Total	YES	NO	Total
2011						
YES	105,812	37,48	143,292	20.69	7.33	28.03
NO	9,598	358,404	368,002	1.88	70.10	71.97
Total	115,41	395,884	511,294	22.57	77.43	100.00

Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

Map 1. HIPGDAC-ES *versus* GEOSTAT2021 - Murcia: Celdas habitadas



Note: For an interactive version of the map, please see https://www.uv.es/goerlich/Ivie/HIPGDAC-ES
Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

To informally examine the discrepancies in absolute terms, it is illustrative to compare both grids, HIPGDAC-ES and GEOSTAT2021, on a map. Maps 1 and 2 give this visual impression for Murcia and Castellón, where, according to table 9, the discrepancies are notable.

In the case of Murcia (**map 1**), the greatest dispersion of the population –in the sense of a greater number of inhabited cells- is evident, especially in the interior and in a more dispersed rural population.

In the case of Castellón (**map 2**), we observe a similar effect, especially in the interior of the province, although somewhat less pronounced. Probably, the existence of buildings classified as residential in Cadastre allows the algorithm of population distribution to assign population to them, although in practice they do not host resident population.

It is possible to offer a quantitative measure of the discrepancy between both grids by a simple modification of the δ discrepancy statistic. In this case both grids add the population of the 2021 census, so that to narrow the index to the interval [0, 1] it is enough to rescale it by the population:

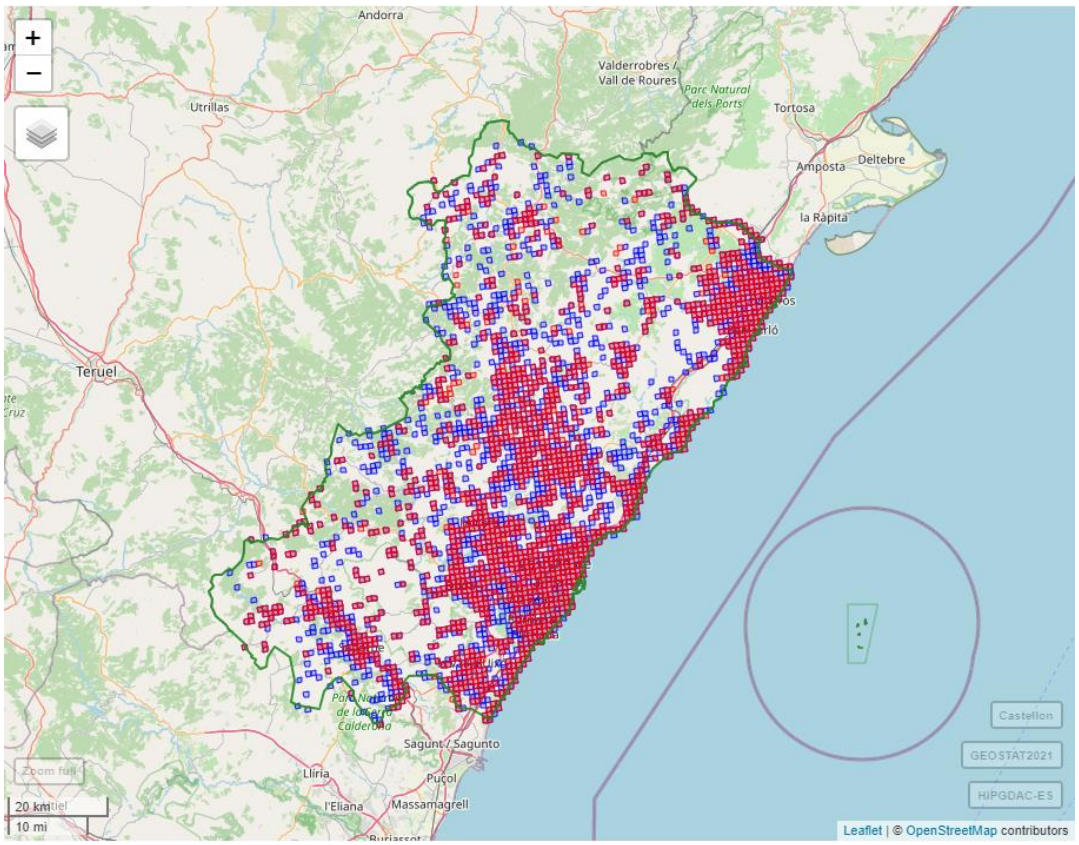
$$\delta' = \frac{\sum_c \left| P_c - P_c^{ref} \right|}{2 \cdot \sum_c P_c} \qquad [2]$$

where **C** indexes the grid cells.

This statistic, δ', ranges between 0, if both grids are identical, and 1, in the case of maximum discrepancy, when there is no overlap between the inhabited cells in both population grids. Thus, and assuming that GEOSTAT2021 is the true distribution of the population, $\text{Prefc in } (2)$, $100 \times \delta'$ can be interpreted as the percentage of population that we place incorrectly on the territory, where the accuracy of the inaccuracy in the location is determined by the grid resolution.

The value of δ', in percentage terms, $100 \times \delta'$, for HIPGDAC-ES and GEOSTAT2021 is 11.8. Which indicates that our disaggregation algorithm places, in 2021, about 12% of the population over the territory incorrectly.

Map 2. HIPGDAC-ES *versus* GEOSTAT2021 - Castellón: Celdas habitadas



Note: For an interactive version of the map, please see https://www.uv.es/goerlich/Ivie/HIPGDAC-ES
Source: HIPGDAC-ES, Census 1900-2021 and INE (GEOSTAT2021).

6.

Concluding comments

This paper presents a general methodology to generate census population grids from conveniently homogenised historical cadastral information **(Uhl *et al.* 2023)** and homogeneous population series at municipal level **(Goerlich, Ruiz, Chorén and Albert 2015)** from 1900 to the last available census. The generated information is available as raster layers in resolutions 100m x 100m and 1km x 1km, and, in the latter case, the grids are also available in vector format.

Although validation of the results is complicated by the lack of historical data, an examination of the trends shown indicates that the results appear reasonable. Comparison with the **GHSL** population grids indicates that our estimates appear more accurate than in these grids, which tend to over-disperse the population, and generate a very high number of inhabited cells. A comparison with the INE 2021 census grid, GEOSTAT2021, yields encouraging results, as they are relatively coincident in many dimensions, although it is clear that our estimates tend to over-disperse the population, especially in rural areas. In fact, this is a well-known bias of the dasymetric methods of spatial disaggregation **(Gallego 2010)**, given the difficulty of limiting the allocation of population to residential areas that are not the main residence for most people. Overall, although the estimates seem far from perfect, they are an improvement on the population distribution from municipal data, where the population is either concentrated in one point –a cell– or spread evenly throughout the municipality.

It should be noted that this work offers initial estimates of grids of census population from 1900 to the present, but that it is an initial estimate –version 0 (*beta*)– that must be submitted to validation by experimentation of the results obtained and, if possible, enhanced with additional available information. Therefore, it should be borne in mind that this is work in progress, that can be improved in the future, and on which any comment is welcome. The development of alternative methods of generating historical population grids, based on the geocoding of gazetteer settlements **(Diez-Minguela, Goerlich and Tirado-Fabregat 2024)**, can shed light on the quality of these estimates.

7.

References

Diez-Minguela, A.; Goerlich, F. J. & Tirado-Fabregat, D. A. (2024) "Nuevos métodos, nueva evidencia sobre el asentamiento de la población en la *Comunitat Valenciana*: la construcción de una *grid* de población para 1887". I Congreso Internacional de Fuentes Geohistóricas: *territorio y sociedad en el tiempo*. Madrid: Centro Cultural UAM 'La Corrala', 23-25 de mayo

Freire S.; MacManus K.; Pesaresi M.; Doxsey-Whitfield E. & Mills J. (2016) "Development of new open and free multi-temporal global population grids at 250m resolution". *Geospatial Data in a Changing World*. Association of Geographic Information Laboratories in Europe (AGILE), AGILE.

Gallego, J. (2010) "A population density grid of the European Union". *Population and Environment*, 31, 460-473.

Goerlich, F. J. (2016) "Una aproximación volumétrica a la desagregación espacial de la población combinando cartografía temática y datos LIDAR". *Revista de Teledetección*, 46 (June): 147-163.

Goerlich, F. J. (2022) "Superficie planimétrica *versus* superficie del paisaje en España. *–superficie 2D* versus *superficie 3D–*" Documento de Trabajo del Ivie. WP-2022-07.

Goerlich, F. J. (2023) *Grid Statistics* –Estadísticas en formato *grid–*. *on-line*. Version: 25/01/2024.

Goerlich, F. J. (2024) *Censo 2021 versus Padrón 2021 –¡Y algunas otras cuestiones censales!–*. Documento de Trabajo del Ivie. WP-2024-01.

Goerlich, F. J. & Cantarino, I. (2014) *Comparing* bottom-up *and* top-down *population density grids: The Spanish Census 2011*. Presented at the 7th European Forum for Geography and Statistics (EFGS) Conference - 22-24 October.

Goerlich, F. J. & Mollá, S. (2021) "Desequilibrios demográficos en España: evolución histórica y situación actual". *Presupuesto y Gasto Público*,102, 31–54.

Goerlich, F. J. & Perez, P. (2021) *LAU2boundaries4spain* version 1.0.0. Part of the rOpenSpain project.

Goerlich, F. J.; Ruiz, F., Chorén, P. & Albert, C. (2015). *Cambios en la Estructura y Localización de la Población. Una visión de largo plazo (1842-2011)*. Bilbao: Fundación BBVA.

Hijmans R. (2022) *terra: Spatial Data Analysis*. *R* package version 1.5-34.

INE (2022) "Censos de Población y Viviendas 2021. Metodología. Versión provisional." Instituto Nacional de Estadística. Subdirección General de Estadísticas Demográficas. Noviembre.

INSPIRE (2014) *D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines* v3.1 INSPIRE Infrastructure for Spatial Information in Europe. European Commission.

Mennis, J. (2003) "Generating surface models of population using dasymetric mapping". *The Professional Geographer*, 55, 31–42.

Pebesma, E. (2018) "Simple Features for R: Standardized Support for Spatial Vector Data". *The R Journal*, 10(1): 439-446. doi: 10.32614/RJ-2018-009.

Pebesma, E. & Bivand, R. (2023). *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

R Core Team (2022) *R: A Language and Environment for Statistical Computing*. Vienna: *R* Foundation for Statistical Computing.

Rappaport, J. & Sachs, J. D. (2003) "The United States as a Coastal Nation". *Journal of Economic Growth*, 8, 5–46. doi: 10.1023/A:1022870216673.

Santos Pérez, L. J. (2019) *Publicación en internet de la documentación gráfica histórica de la Dirección General del Catastro*. Jornadas Ibéricas de Infraestructuras de Datos Espaciales 2019. Cáceres. 23-25 October.

Schiavina, M.; Freire, S.; Carioli, A. & MacManus, K. (2023) *GHS-POP R2023A - GHS population grid multitemporal (1975-2030)*. European Commission, Joint Research Centre (JRC) Dataset.

Schiavina, M.; Melchiorri, M.; Pesaresi, M.; Politis, P.; Freire, S.; Maffenini, L.; Florio, P.; Ehrlich, D.; Goch, K.; Carioli, A.; Uhl, J.; Tommasi, P. & Kemper, T. (2023) *GHSL Data Package 2023*. Public release. GHS P2023. Joint Research Centre. JRC Scientific Information Systems and Databases Report.

Uhl, J. H.; Royé, D.; Burghardt, K.; Aldrey Vázquez, J. A. Borobio Sanchiz, M. & Leyk, S. (2023) "HISDAC-ES: historical settlement data compilation for Spain (1900–2020)". *Earth System Science Data*, 15(October): 4713–4747.

Wickham et al., (2019) "Welcome to the tidyverse". *Journal of Open Source Software*, 4, 43, 1686. doi: 10.21105/joss.01686.

Wright, J. K. (1936) "A method of mapping densities of population". *The Geographical Review*, 26(1): 103–110.


Event
Event Budget for Event : EXPENSES
Actual Cost Breakdown
Estimated vs. Actual
Site
Publicity
Refreshments
Prizes
Program
Miscellaneous
Decorations
$0.00
$10.00
$20.00
$30.00
$40.00
$50.00
$60.00
$70.00
$80.00
$90.00
$100.00

Ivie